

CLAIMS

What is claimed is:

- [c1] 1. A computer-implemented method for using predefined parsing information to analyze web site navigation data in order to identify occurrences corresponding to defined category types, the method comprising:
- for each of multiple distinct web sites each having multiple web pages,
 - receiving web site navigation data associated with the web site that has multiple entries each containing information about a request for a web page of the web site and a response to the request, the request including an indicated URL and sent to a web site server for the web site, each of the indicated URLs optionally including a URL path portion and optionally including a query string portion, the contained information about each request including any URL path portion that is included in the URL indicated for the request and including any query string portion that is included in the URL indicated for the request;
 - retrieving predefined parsing information associated with the web site that includes multiple distinct definitions of logical sites and multiple distinct definitions of category types, each logical site definition specifying an IP address and port number used by a web site server to provide at least some of the web pages of the web site, and each category type definition specifying one of the logical sites and indicating multiple web pages of the web site of that category type by including a URL pattern that is capable of matching the URL paths of the URLs corresponding to the multiple web pages and a query string pattern capable of matching the query strings of those corresponding URLs;
 - for each entry of the received web site navigation data,
 - analyzing the information contained in the entry to determine if the web site server to which the request was sent matches any of the

defined logical sites by using the IP address and the port number specified by that logical site; and

if a logical site is determined to match the web site server, further analyzing the information contained in the entry by storing, for at least one of the category types whose definition specifies the matching logical site, an indication of an occurrence of a request for a web page of that one category type if the information contained in the entry includes a URL path portion that matches the URL pattern included in that one category type definition and includes a query string portion that matches the query string pattern included in that one category type definition;

receiving a request from an operator of the web site to provide information for the web site about occurrences of requests for web pages of specified category types;

retrieving in response the stored indications of the occurrences of the requests for web pages of the specified category types; and

providing the retrieved information to the operator, so that the operators of the multiple web sites can receive information about occurrences of requests for web pages of category types of interest for their web sites.

[c2] 2. The method of claim 1 wherein the web pages of one of the web sites are inter-linked in a hierarchical manner having multiple hierarchy members at multiple hierarchy levels, and wherein the category types defined for that web site correspond to the multiple hierarchy members.

[c3] 3. The method of claim 1 wherein the URL path portions of the URLs that correspond to the web pages of one of the web sites indicate multiple hierarchy members, and wherein the category types defined for that web site correspond to the multiple hierarchy members.

[c4] 4. The method of claim 1 wherein the query string portions of the URLs that correspond to the web pages of one of the web sites include multiple values for a query string parameter name, and wherein the category types defined for that web site correspond to the multiple values.

[c5] 5. The method of claim 1 wherein each of the category type definitions includes a unique name for that category type, and wherein the unique name for at least one of category types is dynamically generated using information contained in at least one entry corresponding to a request for a web page of that one category type.

[c6] 6. The method of claim 1 wherein the category types for at least one of the web sites are hierarchically structured such that, for each of the category types defined for that one web site, the web pages of the web site that correspond to that category type include the web pages of the web site that correspond to each of the category types at a next lower-level in the hierarchy structure.

[c7] 7. The method of claim 1 wherein at least some of the category type definitions indicate multiple web pages by including multiple page type definitions that each specify a distinct combination of a URL pattern and a query string pattern, and wherein a request for a web page that is indicated with a URL is determined to be an occurrence of a request for a web page of a category type if any URL path for the indicated URL and any query string for the indicated URL match the URL pattern and the query string pattern of any of the multiple page type definitions of that category type.

[c8] 8. The method of claim 1 wherein the receiving of the web site navigation data associated with a web site includes retrieving at least one log file from at least one web site server for the web site, the retrieved log files containing the web site navigation data.

[c9] 9. The method of claim 1 including, before the receiving of the web site navigation data for a web site, generating the parsing information associated with the web site based at least in part on the web site servers that can provide web pages of the web site and on the types of web pages that those web site servers can receive.

[c10] 10. The method of claim 1 wherein the received request from an operator of one of the web sites further specifies effective dates such that the information to be provided is for occurrences of the specified category types that took place during the effective dates, and wherein the stored indications that are retrieved in response are for those occurrences.

[c11] 11. The method of claim 1 wherein the operator of one of the web sites from whom a request is received is at a remote location, and wherein the providing of the retrieved information to the operator includes generating a report that includes the retrieved information and sending the generated report to the remote location for presentation to the operator.

[c12] 12. The method of claim 1 wherein the operators of the multiple web sites are customers, and wherein the analyzing of the web site navigation data entries for the web sites is performed as a service for the customers.

[c13] 13. A computer-implemented method for analyzing interaction data to identify occurrences corresponding to defined groups of related items, the method comprising:

receiving an indication of interaction data that is associated with a content set having multiple items, the interaction data having one or more entries that are each related to an interaction with at least one of the items of the content set;

receiving an indication of at least one communication definition that specifies a manner of communicating content set item interactions;

receiving an indication of multiple category type definitions each specifying a group of related content set items and each associated with one of the communication definitions; and

for each entry of the interaction data,

determining whether the entry matches one of the category type definitions in such a manner that the related interaction for the entry is with a content set item that is a member of the group specified by that category type definition and was communicated in the manner specified by the communication definition associated with that category type definition; and

when it is determined that the entry matches one of the category type definitions, storing an indication of an occurrence of that category type.

[c14] 14. The method of claim 13 wherein the content set is a web site with multiple web pages, and wherein the items of the content set are the web pages.

[c15] 15. The method of claim 13 wherein the content set is a group of multiple related web pages that are a subset of web pages of a web site, and wherein the items of the content set are the multiple related web pages.

[c16] 16. The method of claim 13 wherein the content set is multiple related web sites each having multiple web pages, and wherein the items of the content set are the web pages of the multiple related web sites.

[c17] 17. The method of claim 13 wherein the content set is a service providing multiple features, and wherein the items of the content set are the multiple features.

[c18] 18. The method of claim 13 wherein the content set is an executing program providing various functionalities, and wherein the items of the content set are the various functionalities.

[c19] 19. The method of claim 13 wherein the group of related content items specified for each of the category type definitions is a subset of the items for the content set.

[c20] 20. The method of claim 13 wherein each of the interactions related to the interaction data entries includes specifying a Uniform Resource Indicator.

[c21] 21. The method of claim 13 wherein each of the interactions related to the interaction data entries includes requesting that functionality be provided.

[c22] 22. The method of claim 13 wherein each of the interactions related to the interaction data entries includes supplying information.

[c23] 23. The method of claim 13 wherein the manner of communicating content set interactions specified by each of the communication definitions includes using a specified IP address and port number to communicate information related to an interaction.

[c24] 24. The method of claim 13 wherein the manner of communicating content set interactions specified by each of the communication definitions includes using a specified domain name to communicate information related to an interaction.

[c25] 25. The method of claim 13 wherein the manner of communicating content set interactions specified by each of the communication definitions

includes using a specified group of communication parameters to communicate information related to an interaction.

[c26] 26. The method of claim 13 wherein the manner of communicating content set interactions specified by each of the communication definitions includes identifying a specified portion of the content set to which an interaction is to be communicated.

[c27] 27. The method of claim 13 wherein the manner of communicating content set interactions specified by each of the communication definitions includes identifying a specified computing device or computer program provider to which an interaction is to be communicated.

[c28] 28. The method of claim 13 wherein the content set items are stored in multiple directories, and wherein the groups of related content set items specified for the category type definitions are the content set items stored in each of the multiple directories.

[c29] 29. The method of claim 13 wherein the content set is a web site with a home web page having links each corresponding to groups of web pages of the web site, and wherein the groups of related content set items specified for the category type definitions are the groups of web pages.

[c30] 30. The method of claim 13 wherein each of the content set items has an associated URL with a path portion that can include one or more hierarchical members, and wherein each group of related content set items specified for a category type definition includes content set items having a common hierarchical member in the path portion of the URL associated with the content set item.

[c31] 31. The method of claim 13 wherein each of the content set items has an associated URL with a query string portion that includes a common query parameter name and corresponding value, and wherein each group of related content set items specified for a category type definition includes content set items having a common corresponding value for the common query parameter name.

[c32] 32. The method of claim 13 wherein the content set items are each associated with a type of product, and wherein the content set items in each group are related based on the product types associated with those content set items.

[c33] 33. The method of claim 13 wherein the content set items are each associated with one or more features, and wherein the content set items in each group are related based on the features associated with those content set items.

[c34] 34. The method of claim 13 wherein the items of the content set are organized into a hierarchical structure having multiple hierarchy levels and at least one hierarchy member at each hierarchy level, each of the content set items associated with one of the hierarchy members, and wherein each category type corresponds to one of the multiple hierarchy members such that the group of related content set items for that category type includes the content set items associated with that one hierarchy member.

[c35] 35. The method of claim 34 wherein each group of related content set items for a category type corresponding to a hierarchy member further includes the content set items associated with each of the hierarchy members below that hierarchy member in the hierarchical structure.

[c36] 36. The method of claim 34 wherein the hierarchical structure of the content set items is based on a hierarchical manner in which the content set items are stored.

[c37] 37. The method of claim 34 wherein each of the content set items has an associated URL with a path portion, and wherein the hierarchical structure of the content set items is based on a hierarchical structure of the path portions of the associated URLs.

[c38] 38. The method of claim 34 wherein each of the content set items has an associated URL with a query string portion that includes at least one query parameter name and corresponding value, and wherein the hierarchical structure of the content set items is based on the values in the query string portions.

[c39] 39. The method of claim 34 wherein the content set is a web site with a home web page and with multiple other web pages accessible either directly from the home web page or indirectly from the home web page via one or more intervening other web pages, the hierarchical structure such that each of the web pages is a hierarchy member and each of the hierarchy levels includes web pages accessible from the home web page via a same number of other intervening web pages.

[c40] 40. The method of claim 13 wherein each of the interactions related to the interaction data entries includes specifying a Uniform Resource Indicator, wherein each of the content set items has an associated Uniform Resource Indicator, and wherein each of the category type definitions includes a pattern capable of matching at least one Uniform Resource Indicator, the group of related content set items for a category type definition being the content set items whose associated Uniform Resource Indicator matches the included pattern for that category type definition.

[c41] 41. The method of claim 13 wherein each of the interactions related to the interaction data entries includes specifying a URL corresponding to a content set item, the specified URL optionally having a path portion and optionally having a query string portion, each of the query string portions including one or more combinations each having a query parameter name and corresponding query value, and wherein each of the category type definitions includes a URL path pattern capable of matching one or more URL paths and includes a query string pattern capable of matching at least one query string, the group of related content set items for a category type definition being the content set items having corresponding URLs that match the included URL path pattern and query string pattern for that category type definition.

[c42] 42. The method of claim 41 wherein the query string patterns each indicate one or more query parameter names whose presence in a query string is required, allowed, or disallowed if that query string is to match the query string pattern, and wherein the determination of whether a query string portion matches a query string pattern further includes determining if the query string portion includes each of the query parameter names whose presence is indicated in the query string pattern to be required and does not include any of the query parameter names whose presence is indicated in the query string pattern to be disallowed.

[c43] 43. The method of claim 41 wherein at least some of the category type definitions include multiple page type patterns that each include a URL path pattern and a query string pattern, and wherein the related interaction for an entry is determined to be with a content set item that is a member of the group specified by a category type definition if the path portion and the query string portion of the URL corresponding to that content set item match the URL path pattern and the query string pattern of any of the page type patterns for that category type definition.

[c44] 44. The method of claim 41 wherein at least some of the URL path patterns include a static portion capable of matching a single corresponding portion of a URL path and include a variable portion capable of matching multiple corresponding portions of URL paths.

[c45] 45. The method of claim 41 wherein each of the URL path patterns can be specified to match any URL path and wherein each of the query string patterns can be specified to match any query string.

[c46] 46. The method of claim 13 wherein each of the category type definitions includes a unique name for that category type, and wherein the unique name for at least one of category types is dynamically generated using information from at least one interaction with a content set item that is a member of the group specified by that category type definition.

[c47] 47. The method of claim 46 wherein each of the interactions related to the interaction data entries includes specifying a URL corresponding to a content set item, the specified URL having a path portion and a query string portion, the query string portion including at least one query parameter name and corresponding query value, and wherein the information used to dynamically generate the unique name is at least one query value for the specified URL corresponding to the content set item that is a member of the group specified by that category type definition.

[c48] 48. The method of claim 13 wherein at least some of the groups of related content set items contain a single content set item.

[c49] 49. The method of claim 13 wherein each of the communication definitions can be specified to match any manner of communicating content set interactions.

[c50] 50. The method of claim 13 wherein each of the category type definitions can be specified to include any content set item.

[c51] 51. The method of claim 13 wherein at least some of the entries match multiple of the category type definitions.

[c52] 52. The method of claim 51 wherein an indication of an occurrence is stored for each of the multiple category type definitions.

[c53] 53. The method of claim 51 wherein a single indication of an occurrence is stored for the one of the multiple category type definitions having a highest degree of match to the entry.

[c54] 54. The method of claim 51 wherein the category type definitions are hierarchically structured such that the group of related content set items for a category type definition includes the content set items in the groups of related content set items for each of the category type definitions below that category type definition in the hierarchical structure, and wherein a single indication of an occurrence is stored for the one of the multiple category type definitions that is lowest in the hierarchical structure.

[c55] 55. The method of claim 51 wherein the category type definitions are hierarchically structured such that the group of related content set items for a category type definition includes the content set items in the groups of related content set items for each of the category type definitions below that category type definition in the hierarchical structure, and wherein a single indication of an occurrence is stored for the one of the multiple category type definitions that is highest in the hierarchical structure.

[c56] 56. The method of claim 13 wherein each of the entries contain information related to the interaction for the entry, and wherein the determining that an entry matches a category type definition includes analyzing the information contained in the entry.

[c57] 57. The method of claim 13 including receiving a request to provide information about occurrences of specified category types, and providing in response the stored indications of occurrences related to the specified category types.

[c58] 58. The method of claim 13 wherein the determining of whether the interaction data entries match category type definitions is performed as a service for a customer.

[c59] 59. A computer-readable medium whose contents cause a computing device to analyze data to identify occurrences corresponding to defined groups of items, by performing a method comprising:

receiving an indication of data that is associated with a content set having multiple items, the data having one or more entries that are each related to an interaction with at least one of the items of the content set;

receiving an indication of multiple definitions each specifying a group of related content set items; and

for each entry of the data,

determining whether the entry matches one of the definitions in such a manner that the related interaction for the entry is with a content set item that is a member of the group specified by that definition; and

when it is determined that the entry matches one of the definitions, indicating an occurrence of an interaction with the group of items specified by that definition.

[c60] 60. The computer-readable medium of claim 59 wherein the computer-readable medium is a memory of a computer system.

[c61] 61. The computer-readable medium of claim 59 wherein the computer-readable medium is a data transmission medium transmitting a generated data signal containing the contents.

[c62] 62. The computer-readable medium of claim 59 wherein the contents are instructions that when executed cause the computing device to perform the method.

[c63] 63. A computing device for analyzing interaction data to identify occurrences corresponding to defined category types, comprising:

an interaction data receiver component capable of receiving an indication of interaction data that is associated with a content set having multiple items, the interaction data having one or more entries that are each related to an interaction with at least one of the items of the content set;

a definition receiver component capable of receiving an indication of at least one communication definition that specifies a manner of communicating content set item interactions and of receiving an indication of multiple category type definitions each specifying a group of related content set items and each associated with one of the communication definitions; and

an interaction data parsing component capable of, for each entry of the interaction data, determining whether the entry matches one of the category type definitions in such a manner that the related interaction for the entry is with a content set item that is a member of the group specified by that category type definition and was communicated in the manner specified by the communication definition associated with that category type definition and of storing an indication of an occurrence of a category type when it is determined that an entry matches the definition for that category type.

[c64] 64. The computing device of claim 63 wherein the interaction data receiver component, definition receiver component and interaction data parsing component are executing in memory of the computing device.

[c65] 65. A computer system for analyzing interaction data to identify occurrences corresponding to defined category types, comprising:

means for receiving an indication of interaction data that is associated with a content set having multiple items, the interaction data having one or more entries that are each related to an interaction with at least one of the items of the content set;

means for receiving an indication of at least one communication definition that specifies a manner of communicating content set item interactions and for receiving an indication of multiple category type definitions each specifying a group of related content set items and each associated with one of the communication definitions; and

means for, for each entry of the interaction data, determining whether the entry matches one of the category type definitions in such a manner that the related interaction for the entry is with a content set item that is a member of the group specified by that category type definition and was communicated in the manner specified by the communication definition associated with that category type definition, and for storing an indication of an occurrence of a category type when it is determined that an entry matches the definition for that category type.

[c66] 66. A computer-implemented method for analyzing interaction data for a web site to identify occurrences corresponding to defined category types, the method comprising:

receiving an indication of multiple interaction data entries each containing information about an interaction with a web site that includes a specified URL corresponding to one of multiple web pages of the web site, each

of the specified URLs optionally including a URL path portion and optionally including a query string portion, the contained information for each entry including any URL path portion that is included in the specified URL for the entry and including any query string portion that is included in the specified URL for the entry;

receiving an indication of multiple category type definitions that each specify a group of web pages related to a category by using a URL path pattern capable of matching at least one URL path related to the category and using a query string pattern capable of matching at least one query string related to the category; and

for each entry,

analyzing the entry to determine whether the entry matches one of the category type definitions by containing information about a specified URL corresponding to a web page that is related to the category for that category type definition, the matching such that the contained information includes a URL path portion that matches the URL path pattern for that one category type definition and includes a query string portion that matches the query string pattern for that one category type definition; and

when it is determined that the entry matches one of the category type definitions, storing an indication of an occurrence of that category type for the web site.

[c67] 67. The method of claim 66 wherein the contained information about each interaction further includes information related to a manner of identifying a web site server with which the interaction occurred, wherein each of the category type definitions is associated with a logical site definition that specifies a manner of identifying a web site server related to the web site, and wherein the determining that an entry matches a category type definition further includes determining that the information contained in the entry that is related to the manner of identifying the web site server matches the manner of identifying a web

site server specified by the logical site definition associated with that category type definition.

[c68] 68. The method of claim 67 wherein the manner of identifying a web site server related to the web site that is specified by each logical site definition includes using a specified IP address and port number to communicate with the web site server.

[c69] 69. The method of claim 66 wherein each of the interactions with a web site that includes a specified URL includes a request for a web page from that web site that corresponds to the specified URL.

[c70] 70. The method of claim 66 wherein each of the interactions with a web site that includes a specified URL includes a sending to a client of a web page from that web site that corresponds to the specified URL.

[c71] 71. The method of claim 66 wherein each of the web pages has an associated URL with a path portion that can include one or more hierarchical members, and wherein each group of web pages specified for a category type definition includes web pages having a common hierarchical member in the path portion of the URL associated with the web page.

[c72] 72. The method of claim 66 wherein each of the web pages has an associated URL with a query string portion that includes a common query parameter name and corresponding value, and wherein each group of related web pages specified for a category type definition includes web pages having a common corresponding value for the common query parameter name.

[c73] 73. The method of claim 66 wherein the web site is organized into a hierarchical structure having multiple hierarchy levels and at least one hierarchy

member at each hierarchy level, each of the web pages associated with one of the hierarchy members, and wherein each category type corresponds to one of the multiple hierarchy members such that the group of web pages for that category type includes the web pages associated with that one hierarchy member.

[c74] 74. The method of claim 73 wherein each group of web pages for a category type corresponding to a hierarchy member further includes the web pages associated with each of the hierarchy members below that hierarchy member in the hierarchical structure.

[c75] 75. The method of claim 73 wherein the web site has a home web page such that the other web pages are accessible either directly from the home web page or indirectly from the home web page via one or more intervening other web pages, the hierarchical structure such that each of the web pages is a hierarchy member and each of the hierarchy levels includes web pages accessible from the home web page via a same number of other intervening web pages.

[c76] 76. The method of claim 66 wherein each of the URL path patterns can be specified to match any URL path and wherein each of the query string patterns can be specified to match any query string.

[c77] 77. The method of claim 66 wherein at least some of the groups of web pages contain a single web page.

[c78] 78. The method of claim 66 wherein at least some of the entries match multiple of the category type definitions.

[c79] 79. The method of claim 78 wherein an indication of an occurrence is stored for each of the multiple category type definitions.

[c80] 80. The method of claim 66 including receiving a request to provide information about occurrences of specified category types, and providing in response the stored indications of occurrences related to the specified category types.

[c81] 81. A computer-readable medium containing instructions that when executed cause a computer system to analyze data to identify occurrences corresponding to defined groups of web pages, by performing a method comprising:

receiving an indication of multiple data entries each containing information about an interaction with a web site that includes a specified URL corresponding to one of multiple web pages of the web site, each of the specified URLs optionally including a URL path portion and optionally including a query string portion, the contained information for each entry including any URL path portion that is included in the specified URL for the entry and including any query string portion that is included in the specified URL for the entry;

receiving an indication of multiple definitions that each specify a group of web pages related to a category by using a URL path pattern capable of matching at least one URL path related to the category and using a query string pattern capable of matching at least one query string related to the category; and

for each entry,

analyzing the entry to determine whether the entry matches one of the definitions by containing information about a specified URL corresponding to a web page that is related to the category for that definition, the matching such that the contained information includes a URL path portion that matches the URL path pattern for that one definition and includes a query string portion that matches the query string pattern for that one definition; and

when it is determined that the entry matches one of the definitions, indicating an occurrence of an interaction with the group of web pages specified by that definition.

[c82]

82. A computer-implemented method for analyzing interaction data for a web site to identify occurrences corresponding to defined category types, the method comprising:

receiving an indication of multiple interaction data entries each containing information about a request that specifies a URL corresponding to a web page of a web site, each of the specified URLs optionally including a URL path portion and optionally including a query string portion, each of the query string portions including one or more combinations each having a query parameter name and corresponding query value, the contained information about each request including any URL path portion that is included in the specified URL for the request and including any query string portion that is included in the specified URL for the request;

receiving an indication of a category type definition corresponding to multiple categories, the category type definition specifying a URL path pattern capable of matching at least one URL path related to the multiple categories and a query string pattern capable of matching at least one query string related to the multiple categories, each query string pattern indicating one or more query parameter names, the category type definition further specifying a name definition for providing a unique name for each of the multiple categories, the name definition including at least one of the indicated query parameter names and indicating how values for each of the included query parameter names are to be combined to form the names of the multiple categories, each unique combination of values for the indicated query parameter names corresponding to one of the multiple categories; and

for each entry,

analyzing the entry to determine whether the entry matches one of the category type definitions by containing information about a request corresponding to a web page that is related to the category for that category type definition, the matching such that the information contained in the entry includes a URL path portion that matches the URL path pattern specified in that one

category type definition and includes a query string portion that matches the query string pattern specified in that one category type definition; and

when it is determined that the entry matches one of the category type definitions,

determining the name of the category to which the entry corresponds by retrieving the value from the query string portion of the contained information for the entry that corresponds to each of the query parameter names included in the name definition and by combining the retrieved values in the manner indicated in the name definition; and

storing an indication of an occurrence of the category having the name formed by the combined retrieved values.

[c83] 83. The method of claim 82 wherein the contained information about each request further includes information related to a manner of identifying a web site server to which the request was sent, wherein each of the category type definitions is associated with a logical site definition that specifies a manner of identifying a web site server related to the web site, and wherein the determining that an entry matches a category type definition further includes determining that the information included in the entry that is related to the manner of identifying the web site server matches the manner of identifying a web site server specified by the logical site definition associated with that category type definition.

[c84] 84. A computer-implemented method for analyzing interaction data for a web site to identify occurrences corresponding to defined category types, the method comprising:

receiving an indication of multiple interaction data entries each containing information about a request that specifies a URL corresponding to a web page of a web site, each of the specified URLs optionally including a URL path portion and optionally including a query string portion, the contained information about each request including any URL path portion that is included in

the specified URL for the request and including any query string portion that is included in the specified URL for the request;

receiving an indication of multiple category type definitions that each specify a group of web pages related to a category with multiple page type patterns that each specify a distinct combination of a URL path pattern capable of matching at least one URL path related to the category and a query string pattern capable of matching at least one query string related to the category; and

for each entry,

analyzing the entry to determine whether the entry matches one of the category type definitions by containing information about a request corresponding to a web page that is related to the category for that category type definition, the matching such that, for any of the page type patterns included in that one category type definition, the information contained in the entry includes a URL path portion and a query string portion that match the URL path pattern and the query string pattern specified in that page type pattern; and

when it is determined that the entry matches one of the category type definitions, storing an indication of an occurrence of that category type for the web site.

[c85]

85. The method of claim 84 wherein the contained information about each request further includes information related to a manner of identifying a web site server to which the request was sent, wherein each of the category type definitions is associated with a logical site definition that specifies a manner of identifying a web site server related to the web site, and wherein the determining that an entry matches a category type definition further includes determining that the information included in the entry that is related to the manner of identifying the web site server matches the manner of identifying a web site server specified by the logical site definition associated with that category type definition.

[c86] 86. A computer-implemented method for analyzing interaction data for a web site to identify occurrences corresponding to defined category types, the method comprising:

receiving an indication of multiple interaction data entries each containing information about a request that specifies a URL corresponding to a web page of the web site, each of the specified URLs optionally including a URL path portion and optionally including a query string portion, each of the query string portions including one or more combinations each having a query parameter name and corresponding query value, the contained information about each request including any URL path portion that is included in the specified URL for the request and including any query string portion that is included in the specified URL for the request;

receiving an indication of multiple category type definitions that each specify a group of web pages related to a category with a URL path pattern capable of matching at least one URL path related to the category and a query string pattern capable of matching at least one query string related to the category, each query string pattern indicating one or more query parameter names whose presence in a query string is required, allowed, or disallowed if that query string is to match the query string pattern; and

for each entry,

analyzing the entry to determine whether the entry matches one of the category type definitions by containing information about a request corresponding to a web page that is related to the category for that category type definition, the matching such that the information contained in the entry

(a) includes a URL path portion that matches the URL path pattern specified in that one category type definition and

(b) includes a query string portion that includes each of the query parameter names whose presence is indicated in the query string pattern specified in that one category type definition to be required, and that does not include any of the query parameter names whose presence is indicated in the

query string pattern specified in that one category type definition to be disallowed;
and

when it is determined that the entry matches one of the category type definitions, storing an indication of an occurrence of that one category type for the web site.

[c87] 87. The method of claim 86 wherein the contained information about each request further includes information related to a manner of identifying a web site server to which the request was sent, wherein each of the category type definitions is associated with a logical site definition that specifies a manner of identifying a web site server related to the web site, and wherein the determining that an entry matches a category type definition further includes determining that the information included in the entry that is related to the manner of identifying the web site server matches the manner of identifying a web site server specified by the logical site definition associated with that category type definition.

[c88] 88. The method of claim 86 wherein each of the query string patterns additionally indicates a type of value corresponding to at least some of the indicated query parameter names, and wherein the determining that an entry matches a category type definition further includes determining that, for each of the query parameter names that is included in the query string portion of the contained information for the entry and that is indicated to have a type of value in the query string pattern specified by that one category type definition, the corresponding query value in the query string portion matches the indicated value type.

[c89] 89. A computer-implemented method for analyzing interaction data to identify occurrences corresponding to defined hierarchies of items, the method comprising:

receiving an indication of multiple interaction data entries each containing information related to an interaction with one of multiple items of a content set, the content set items structured in a hierarchy having multiple hierarchy members at multiple hierarchy levels;

receiving an indication of multiple hierarchy definitions that each correspond to one or more related hierarchy members; and

for each entry,

analyzing the entry to determine whether the entry matches one of the hierarchy definitions by containing information about an interaction with a content set item that is one of the hierarchy members to which that one hierarchy definition corresponds; and

when it is determined that the entry matches one of the hierarchy definitions, indicating an occurrence of an interaction with the related hierarchy members to which that one hierarchy definition corresponds.

[c90] 90. The method of claim 89 wherein the content set items are web pages of a web site, wherein each of the interaction data entries contains a URL specified as part of a request corresponding to one of the web pages, each specified URL optionally including a URL path portion and optionally including a query string portion, wherein each of the hierarchy member definitions includes a URL path pattern capable of matching at least one URL path related to the one or more web pages corresponding to the hierarchy member and includes a query string pattern capable of matching at least one query string related to the one or more web pages corresponding to the hierarchy member, and wherein the matching of a hierarchy definition to an entry is based on the URL contained in the entry including a URL path portion that matches the URL path pattern for that

one hierarchy member definition and including a query string portion that matches the query string pattern for that one hierarchy member definition.

[c91] 91. The method of claim 90 wherein the contained information about each request further includes information related to a manner of identifying a web site server to which the request was sent, wherein each of the hierarchy member definitions is associated with a logical site definition that specifies a manner of identifying a web site server related to the web site, and wherein the determining that an entry matches a hierarchy member definition further includes determining that the information included in the entry that is related to the manner of identifying the web site server matches the manner of identifying a web site server specified by the logical site definition associated with that category type definition.

[c92] 92. The method of claim 89 wherein the content set items are structured in the hierarchy based on product types associated with the content set items

[c93] 93. The method of claim 89 wherein the content set items are structured in the hierarchy based on features associated with the content set items.

[c94] 94. The method of claim 89 wherein the content set items are structured in the hierarchy based on categories associated with the content set items.

[c95] 95. A computer-implemented method for analyzing usage data to identify occurrences corresponding to defined groups of features, the method comprising:

receiving an indication of usage data associated with a provided service or an executing computer program that has multiple features available for use, the usage data having multiple entries each related to a distinct use of one of multiple features of the provided service or executing computer program that includes information being communicated;

receiving an indication of multiple definitions each specifying a group of features related to a category and each associated with a manner of communicating information to the provided service or to the executing computer program; and

for each entry of the usage data,

determining whether the entry matches one of the definitions such that the related use for the entry is of a feature that is a member of the group of features specified by that definition and such that the information communicated for the related use is communicated in the manner associated with that definition; and

when it is determined that the entry matches one of the definitions, storing an indication of an occurrence of a use of the group of features specified by that definition.

[c96]

96. A computer-readable medium containing a data structure that stores multiple definitions for category types so that occurrences of those category types in interaction data for a web site can be identified, the data structure having multiple entries each corresponding to a category type definition that specifies a group of web pages related to a category, each entry storing a URL path pattern capable of matching at least one URL path related to the category and a query string pattern capable of matching at least one query string related to the category,

such that when analyzing information about an interaction with a web page of the web site having a specified URL that optionally includes a URL path portion and optionally includes a query string portion, if the web page is determined to be a member of the group of web pages specified by a category type definition then an occurrence of that category type is indicated, the web page determined to be a member of the group of web pages for a category type definition if the specified URL includes a URL path portion that matches the URL path pattern specified for that category type definition and includes a query string

portion that matches the query string pattern specified for that category type definition.

[c97] 97. The computer-readable medium of claim 96 wherein each of the entries further includes an indication of a logical site definition that specifies a manner of identifying a web site server related to the web site,

such that, when the information about the interaction further includes information related to a manner of identifying a web site server with which the interaction occurred, the web page is determined to be a member of the group of web pages for a category type definition only if the information related to the manner of identifying the web site server matches the manner of identifying a web site server specified by the logical site definition indicated by that category type definition.

[c98] 98. The computer-readable medium of claim 96 wherein the category type definitions are related to each other and wherein at least some of the entries further include an indication of a relationship of the category type definition for that entry to at least one other category type definition,

such that, when the web page is determined to be a member of the group of web pages for a category type definition whose entry includes an indication of a relationship to other category type definitions, the web page is also determined to be a member of the group of web pages for at least some of the other related category type definitions.

[c99] 99. The computer-readable medium of claim 96 wherein the category type definitions corresponding to at least some of the entries each have multiple distinct combinations of a URL path pattern and a query string pattern, the entry for each of those category type definitions further storing the multiple combinations of URL path patterns and query string patterns of the patterns for that category type definition,

such that the web page is determined to be a member of the group of web pages for a category type definition having multiple combinations if, for any of those combinations, the information includes a URL path portion that matches the URL path pattern specified in that combination and includes a query string portion that matches the query string pattern specified in that combination.

[c100] 100. The computer-readable medium of claim 96 wherein the stored query string patterns each indicate one or more query parameter names whose presence in a query string is required, allowed, or disallowed if that query string is to match the query string pattern,

such that a query string portion of the information is determined to match the query string pattern specified for one of the category type definitions if the query string portion includes each of the query parameter names whose presence is indicated in that query string pattern to be required and does not include any of the query parameter names whose presence is indicated in that query string pattern to be disallowed.

[c101] 101. The computer-readable medium of claim 96 further containing a data structure having multiple entries that each store an exclusion definition that specifies a type of interaction,

such that if the information being analyzed is of a type matching one of the exclusion definitions, the information will not be determined to match any of the category type definitions.

[c102] 102. A computer-readable medium containing a data structure storing multiple definitions for category types so that occurrences of those category types can be identified in interaction data or usage data, the data structure having multiple entries each corresponding to a category type definition, each entry specifying a group of related content set items for a content set and including an

indication of a communication definition that specifies a manner of communicating information related to interactions or uses corresponding to the content set items,

so that when analyzing data about an interaction or use that corresponds to a content set item and that indicates a manner in which related information was communicated, if the data matches one of the category type definitions in such a manner that the interaction or use corresponds to one of the content set items in the group specified by that category type definition and had related information that was communicated in the manner specified by the communication definition indicated by that category type definition, an occurrence of that category type can be identified.

[c103] 103. A method for analyzing customer data to identify occurrences corresponding to defined categories, the method comprising:

receiving a request from a customer to analyze interaction or usage data for that customer related to a content set having multiple content set items;

receiving an indication of definitions for the customer that each specify a group of content set items related to a category and are each associated with at least one manner of communicating information;

receiving a first set of data for the customer that includes information about at least one interaction or use;

analyzing the received set of data to determine whether the received data includes information about any interactions or uses that match one of the definitions in such a manner that the interaction or use is with a content set item in the group specified by that definition and had related information communicated in a manner associated with that definition; and

when it is determined that the received data matches one of the definitions, providing information to the customer about an occurrence for that category.

[c104] 104. The method of claim 103 wherein sets of data are automatically retrieved from the customer and analyzed on a periodic basis.

[c105] 105. The method of claim 103 wherein the providing of the information to the customer includes generating reports on a periodic basis and sending the generated reports to the customer.

[c106] 106. The method of claim 103 including storing an indication of the occurrence for that category, and wherein the providing of the information to the customer includes receiving a request from the customer at a remote location to provide information about occurrences of one or more categories and sending the requested information to the remote location.

[c107] 107. The method of claim 103 wherein the method is performed for multiple customers each having distinct interaction or usage data and having distinct definitions.

[c108] 108. A method for creating definitions of category types for analyzing interaction data for a web site to identify occurrences corresponding to defined category types, the method comprising:

receiving an indication of a log file for the web site or of other information related to the web site that indicates multiple interactions with a web site server for the web site, each indicated interaction having associated information including network address information for the web site server and a URL specified as part of the interaction;

analyzing the log file or the other information to identify distinct network addresses for the web site servers for the web site, and generating a site definition for each of the identified network addresses that includes that network address; and

analyzing the log file or the other information to identify groups of related web pages, and generating a category type definition for each of the identified groups,

so that information about an interaction with the web site can be analyzed to determine whether the information matches one of the category type definitions in such a manner that the interaction is with a web page that is a member of the group specified by that category type definition and was with a web site server having a network address that matches one of the site definitions.

[c109] 109. The method of claim 108 wherein the specified URLs each optionally include a URL path portion and optionally include a query string portion, and wherein the category type definitions are each generated to include a URL path pattern capable of matching at least one URL path related to the group of web pages for that category type definition and a query string pattern capable of matching at least one query string related to the group of web pages for that category type definition.

[c110] 110. The method of claim 108 wherein each of the category type definitions are further generated to include an indication to one of the generated site definitions, such that information about an interaction with the web site is not determined to match one of the category type definitions unless the interaction was with the web site server having the network address included in the site definition indicated by that one category type definition.

[c111] 111. The method of claim 108 including analyzing information about an interaction with the web site using the generated category type definitions and site definitions.